Systematic deletions in *Bacillus subtilis* genome coupled with large-scale metabolic model refinement

Kosei Tanaka[1,2], Christopher S. Henry[3,*], Jenifer F. Zinner[3], Edmond Jolivet[1], Matthew P. Cohoon[3], Fangfang Xia[3], Vladimir Bidnenko[1,2], S. Dusko Ehrlich[1,2], Rick L. Stevens[3,4], and Philippe Noirot[1,2,*]

[1] INRA, UMR 1319 Micalis, Jouy-en-Josas F-78350, France.

[2] AgroParisTech, UMR Micalis, Jouy-en-Josas F-78350, France

[3] Mathematics and Computer Science Department, Argonne National Laboratory, S. Cass Avenue, Argonne, IL 60439, USA.

[4] Computation Institute, The University of Chicago, S. Ellis Avenue, Chicago, IL 60637, USA.

* Correspondence to chenry@mcs.anl.gov and philippe.noirot@jouy.inra.fr

Running title: Systematic deletions in *Bacillus subtilis* genome

Exact character count: 44032

Subject Categories: Synthetic biology, Functional genomics

Abstract

Here we establish a comprehensive catalogue of the regions of the *B. subtilis* genome that are dispensable for growth in a variety of media conditions. In complex medium, we attempted deletion of 157 regions, ranging in size from 2 to 159 kb and covering ~76% of the chromosome. 146 deletions were successful, while the remaining regions were subdivided to identify new essential genes (4) and coessential gene sets (7). We screened for viability of our mutant strains in rich defined medium and glucose minimal media. Experimental observations were compared with predictions by the *i*Bsu1103 model, revealing discrepancies that led to numerous model refinements. We ultimately produced the *i*Bsu1103V2 model and generated predictions of metabolites that could restore the growth of unviable strains. These predictions were experimentally tested and demonstrated to be correct for 27 strains, validating the model refinements. Importantly *i*Bsu1103V2 has improved considerably at predicting loss of viability. Together with the deletion mutant repertoire, our combined modeling/experimental approach opens the way to the model-assisted design of *B. subtilis* strains for biotechnology applications.

Keywords: Chemically defined medium / Essential genes / Flux balance analysis / Growth phenotypes / *one more*

Introduction

Biological systems often maintain phenotypic stability when exposed to diverse perturbations arising from environmental changes, intracellular stochastic events (or noise), and genetic variation. This robustness is an inherent property of all biological systems and is strongly favored by evolution (de Visser *et al*, 2003). The robustness of bacterial networks is well demonstrated through systematic genetic perturbations. For example, the analysis of flux distribution in 137 null mutants of *Bacillus subtilis* showed that the metabolic state of *B. subtilis* under a given condition is extremely stable and robust against random genetic mutations (Fischer and Sauer, 2005). Also, the rewiring of transcriptional regulatory circuits in *Escherichia coli* revealed that the bacterium tolerates potentially radical changes in its circuitry, with limited genome-wide transcriptional changes observed in most cases, indicating that bacterial networks have a built-in resilience to change (Isalan *et al*, 2008). Functional robustness arises from the many redundancies, interlocking pathways and feedback mechanisms inherent to the complexity of biological networks (Stelling *et al*, 2004). This complexity enables the biological systems to dynamically adapt or compensate for losses or environmental changes. However, it also makes biological systems resistant to re-engineering and could be problematic for the construction and operation of engineered genetic circuits designed to create or modify biological functions. With the prospect of synthetic biology creating genetic circuits of increasing size (Purnick and Weiss, 2009; Serrano, 2007), the complexity of cells might hinder the practical applications of synthetic approaches.

A strategy to minimize cellular complexity is to study the simplest bacteria capable of growth in laboratory conditions, the mycoplasmas. The underlying assumption is that a simple cell includes less redundancy among genes and regulatory circuits and will be easier to understand and to re-engineer. The tiny *Mycoplasma* genome was used to identify the minimal set of genes sufficient to sustain

bacterial life (Glass *et al*, 2006), and was chemically synthesized, assembled and transplanted into a bacterial cell whose growth was controlled by the synthetic chromosome (Gibson *et al*, 2010). This technology, in combination with the new knowledge brought by large-scale systems biology analyses of a closely related *Mycoplasma* species (Guell *et al*, 2009; Kuhner *et al*, 2009; Yus *et al*, 2009), opens the possibility that in the near future the functions of all the genes essential for the life of a near-minimal bacterium will be understood at some level (Glass *et al*, 2009) (Henry *et al*, 2010).

Another strategy to reduce cell complexity is to simplify the genomes of well studied bacteria through the removal of nonessential genetic elements. Large-scale knockout approaches have been undertaken in model bacteria heavily used in biotechnology such as *Escherichia coli* (Hashimoto *et al*, 2005; Kolisnychenko *et al*, 2002; Posfai *et al*, 2006; Yu *et al*, 2002) and *Bacillus subtilis* (Morimoto *et al*, 2008; Westers *et al*, 2003). A key objective in most of these studies was to eliminate genes unnecessary for growth without compromising the cell physiology in order to reduce the metabolic burden of the cell and possibly to improve its metabolic efficiency. To this end, the chromosome regions targeted for deletion were mostly selected among mobile DNA elements, cryptic prophages and virulence genes, genes required for growth in specific environments, and genes of unknown functions. The stepwise accumulation of deletions yielded *E. coli* with a genome reduced by 14.3%, which was shown to be a better host for the propagation of recombinant genes and to have enhanced electroporation efficiency (Posfai *et al*, 2006). This reduced strain was engineered for increased productivity of L-threonine (Lee *et al*, 2009). Similarly, a *B. subtilis* strain with a genome reduced by 20% was shown to have an increased productivity for recombinant proteins (Morimoto *et al*, 2008). However, the most reduced *E. coli* strain (by 29.7%) exhibited severely impaired chromosome organization (Hashimoto *et al*, 2005). Together, these studies indicate that genome reduction has a high potential for biotechnological applications and the extent of genome reduction is limited not only by the lack of knowledge about important cell cycle

functions but also by the lack of a comprehensive repertoire of the nonessential chromosome regions. From such a repertoire, a much larger number of deletion combinations could be attempted, potentially leading to further genome reduction while preserving cell physiology.

Here, we report the systematic mapping of the nonessential regions in the 4.21 MB chromosome of *B. subtilis 168*. We developed unbiased criteria for the selection of chromosome regions that need to be preserved to retain viability and natural competence, which in turn delineated 157 chromosome intervals that were systematically targeted for deletion. Interval deletions were attempted on two rich media; the complex LB medium and the NMS medium of chemically defined composition. Deletion mutants were obtained for 146 intervals on one or more media conditions, while 10 intervals could not be deleted in any condition. We employed the *i*Bsu1103 model (Henry *et al*, 2009) to predict interval deletion outcomes on various media and identify potential metabolic causes for loss of viability. The discrepancies between predictions and observations were further investigated experimentally, and where possible, the model was corrected to improve accuracy. Overall, this approach identified new essential and coessential genes, and improved our understanding of *B. subtilis* metabolism and growth. We report the phenotypes of all interval strains tested, the model predictions generated to explain metabolic phenotypes observed, and a refined version of the iBsu1103 model (iBsu1103V2) that integrates all corrections made in response to observed phenotypes.

Results

Systematic deletion of chromosome intervals

The chromosome regions targeted for deletion were defined relative to the 271 genes reported to be individually essential for *B. subtilis* survival in Luria-Bertani (LB) medium at 37°C (Kobayashi *et al*, 2003). In addition, 254 genes involved in cellular processes essential for the experimental procedures, such as DNA recombination and repair, SOS response, competence development and transformation, and pyrimidine salvage pathway were also preserved. To avoid potential polar effects that could alter the expression of the preserved genes, the genes lying in the same operon as a preserved gene were also kept. Using the information on *B. subtilis* operons available in DBTBS (Sierro *et al*, 2008), 80 and 51 genes were preserved based on experimentally identified and computationally predicted operons, respectively. Together, these 656 preserved genes defined the chromosome intervals to be deleted (Supplementary information 1). However, upon manual curation of intervals, we realized that the positions of some interval endpoints could potentially inactivate the promoters of preserved genes due to a lack of knowledge about the transcriptional start sites. Also, the structure of the chromosome after deletion was taken into consideration to avoid the head-on collisions of transcription units, which could be potentially deleterious for the strain (Supplementary Figure S1). As a result, 157 additional genes were preserved. Altogether, a list of 813 preserved genes was established (Supplementary Table S1), defining 172 intervals to be deleted in the *B. subtilis* chromosome. Fifteen intervals consisting of a single gene previously reported to be dispensable (Kobayashi *et al*, 2003) were excluded from the list, leaving 157 candidate intervals to be deleted (Figure 1, Supplementary Table S2).

Each chromosome interval was deleted in the master strain (see Material and Methods) by the homologous replacement of the interval with a cassette carrying the selectable phleomycin-resistance

gene (*phleo*) (Figure 2A). For each interval, 2 primer pairs (p1-p2, p3-p4) were computationally designed (Supplementary information 2) and used to PCR-amplify the DNA segments flanking the interval to be deleted, using the master strain chromosomal DNA as a template (Figure 2B). These segments were joined with the *phleo* cassette in a subsequent PCR reaction (Fabret *et al*, 2002), and the assembled DNA molecule was used to transform competent cells of the master strain. Assembled DNA molecules were obtained for all intervals except i0031 for which the adjacent tRNA genes prevented PCR amplification due to the formation of inhibitory higher-order DNA structures. Transformations were performed at 37°C on NMS, a rich medium of chemically defined composition designed for this study (see Material and Methods). The replacement of the targeted chromosome interval by the cassette yielded phleomycin-resistant transformants, which were purified twice by single colony isolation and were assayed for the presence of the integrated cassette and for the absence of the targeted chromosome interval (Figure 2C). Overall, 135 deletion mutants were obtained on selective NMS plates (Figure 1, Supplementary Table S2). Of note, two deletions (intervals i0308 and i0606) could not be obtained with the *phleo* marker but were constructed using a chloramphenicol-resistance (*cat*) marker. This marker-specific effect was caused by a much greater sensitivity to phleomycin of the deletion mutants relative to the master strain. The qualitative assessment of colony growth (Figure 2D) indicated that 8 deletions conferred a slow growth phenotype. The 135 deletion mutant strains grew at 37°C on LB, a rich complex medium of ill defined composition, indicating that the functions encoded by each interval are dispensable for cell survival under the 2 conditions tested.

In contrast, 21 chromosome intervals could not be deleted after selection on NMS plates (Figure 1, Supplementary Table S2). However, 11 deletion mutants were obtained on selective LB medium, suggesting that the intervals encode functions essential for growth on NMS and dispensable for growth on LB. The remaining 10 intervals could not be deleted under any condition using the *phleo* and *cat*

selection markers, suggesting that they encode functions essential for growth (Figure 1, Supplementary Table S2). Altogether, 146 deletion mutants were obtained in at least one condition with the cumulated size of dispensable intervals covering 3.05 MB.

Quantitative measurements of the growth rates of 137 deletion mutants that formed normal colonies on LB and of 3 deletion mutants that exhibited slow colony growth on LB (Supplementary Table S2) were performed in NMS-containing microtiter plates at 37°C under mild aeration (Supplementary information 3). Under this condition, the maximal doubling times were distributed in 3 subgroups (Supplementary Figure S3): 110 deletion mutants formed a homogeneous subgroup in which doubling times exhibited a bell-shaped distribution centered on an average doubling time (56.6 ± 7.7 min) very similar to that of the master strain (57.1 ± 8.4 min); 23 deletion mutants formed the tail of the distribution and exhibited slow growth (d > 75 min), and 7 mutant strains for which no growth was detected (Supplementary Table S2). Importantly, these results were in good agreement with the qualitative phenotypes of colony growth (Supplementary information 3). Altogether, these findings indicate that deletion of most dispensable intervals have either a small positive or negative effect on cell growth in NMS. However, some deletions severely affected growth, suggesting that the corresponding intervals encode functions required for normal growth under the conditions tested.

Identification of essential functions via the splitting of intervals

The 21 chromosome intervals encoding functions essential for growth on NMS include between 2 and 52 genes (Supplementary Table S2). Among the 11 deletion mutants obtained on LB but unable to grow on NMS, all but one (Δ0663) recovered the capacity to form colonies on NMS supplemented with 10% LB. Since 10% LB did not by itself support colony formation, it likely provided metabolites required for growth of the deletion mutants on NMS. Thus, to identify these metabolic functions and further validate

and test the genome-scale model, the intervals larger than 4 genes were split into smaller intervals which were individually deleted, and the capacity of the deletion mutant strains to grow on NMS and LB was tested. This cycle was repeated up to 3 times until the function required for growth could be narrowed down to a few genes (Figure 1, Supplementary Table S3). As a result, the 21 intervals were subdivided into smaller intervals containing from 1 to 12 genes that encoded functions required for growth on NMS. This approach identified: i) 4 genes individually essential on LB (*patA*, *ribC*, *rnpB*, and *ywpB*) that were not discovered in a previous systematic gene knockout study (Kobayashi *et al*, 2003); ii) 5 new co-essential gene pairs *glnR – glnA (i0787-3)*, *ywfI (hemQ) – ywfH (i0825)*, *yneE (sirA)-yneF (i0883)*, *ccpA – aroA (i0915)*, *tyrA – hisC (i0832)*; and iii) 2 essential intervals of 8 (i0797) and 10 genes (i0833) which could not be reduced further. More detailed information on the essential functions encoded by these intervals is provided (Supplementary information 4).

Modeling of deletion intervals using the *i*Bsu1103 metabolic model of *B. subtilis*

Metabolic modeling was applied to predict the viability of all 157 deletion mutants in both NMS and LB media conditions. We performed this modeling to explore the accuracy of the model by testing for agreement between observed and predicted viability and to produce testable hypotheses to explain the loss of viability of some strains in LB and/or NMS media. All metabolic modeling was conducted using the *i*Bsu1103 genome-scale model of *B. subtilis* 168 (Henry *et al*, 2009), with adjustments to account for the mutations inactivating the *trpC* and *trpD* genes which are present in the master strain used as a recipient for all genetic transformations (Materials and Methods). We call this adjusted model *i*Bsu1103ΔtrpCD.

Of the 135 deletion mutants that grew in both LB and NMS media, the model correctly predicted viability for 130 (96%) strains. In the five remaining cases, the model falsely predicted that the strains

would be unviable in both LB and NMS media (Supplementary Table S2). Of the 11 strains that grew in LB but not NMS, the model correctly predicted viability for 3 (27%) of them. For the 8 remaining strains, the model incorrectly predicted growth (7) or lack of growth (1) in both LB and NMS. The low success rate for this class of observed phenotype indicated an inability by the initial model and media formulations to differentiate between the LB and NMS conditions. Finally, of the 10 deletion strains that could not be obtained on NMS or LB, the model made one correct prediction, two partly correct predictions for lack of growth on NMS, and 7 false predictions for growth in both LB and NMS (Supplementary Table S2). Overall, these results indicate a poor capacity by the original model to predict loss of viability. This was anticipated because the model covers metabolic subsystems only and cannot capture the phenotypes arising from the disruption of nonmetabolic systems. Also, the information on new essential functions was not included in the model except for *ywpB* (*fabZ*) required for fatty acid biosynthesis. Interestingly, the model analysis, our experimental evidence (supplementing NMS with 10% LB, see above), and our survey of existing knowledge suggested that most of the observed discrepancies (false-positive predictions) are of metabolic origin.

Christopher 1/10/12 11:37 PM
Deleted: 26

Modeling and experimentally testing growth of deletion mutants on minimal media

To further validate our genome-scale metabolic model, we also tested viability of our deletion strains in glucose minimal media, using either ammonium (MM+NH4) or glutamine (MM+Gln) as sole nitrogen sources. Unlike LB and NMS media, these minimal media conditions (MM) require the cell to utilize metabolic pathways to produce all biomass precursor compounds from a small number of simple substrates. In this way, we maximize the use of biological components that are included in the metabolic model, improving the utility of our data for model validation. In these experiments, cells were deposited on plates of MM+NH4 or MM+Gln and growth was monitored as described in Material and

methods (Supplementary Table S2). As expected, the seven deletion mutants that grew in LB but not in NMS media also failed to grow in MM. Eighteen deletion mutants growing in NMS but not in MM were identified. Interestingly, 5 deletion mutants (Δ0281, Δ0581, Δ0676, Δ0704, and Δ0708) grew only on MM+Gln, suggesting that these strains are deficient in their capacity to utilize ammonium as a nitrogen source. One other deletion mutant (Δ0641) grew only on MM+NH4, suggesting that this strain is deficient for the utilization of glutamine as a nitrogen source (Supplementary Table S2).

We applied the *i*Bsu1103ΔtrpCD model to predict the viability of all strains in MM+NH4 and MM+Gln media (Supplementary Table S2). Of the 109 strains that grew in both MM, 106 (97%) were correctly predicted by the model. The remaining three strains were falsely predicted not to grow in either minimal media. Of the 18 deletion strains growing in NMS but not in MM, 6 (33%) were correctly predicted by the model. For the remaining strains, the model falsely predicted growth (11) in both MM or no growth (1) in NMS and MM. Two of the 5 strains that grew in MM+Gln but not MM+NH4 were correctly predicted, while the remaining three were predicted to grow in both MM. Similarly, the only strain that grew in MM+NH4 but not MM+Gln was incorrectly predicted to grow in both MM.

In summary, the *i*Bsu1103ΔtrpCD model performed well at predicting growth phenotypes on all tested media with an overall accuracy of 96% (Table 1). However, it performed poorly with an overall accuracy of 36% (Table 1) at predicting lack of growth and at distinguishing between the two minimal media and between the two rich media.

Refining the *i*Bsu1103 model and annotations to improve accuracy

All model prediction errors fall into two categories: (i) *false positive* predictions when a strain observed to be unviable is predicted to be viable; and (ii) *false negative* predictions when a strain observed to be viable is predicted to be unviable. Our studies with the initial *i*Bsu1103ΔtrpCD model indicate that *false*

positive prediction errors (occurring for 64% of unviable strains) are far more common than *false negative* prediction errors (occurring for only 4% of viable strains) (Table 1). We analyzed the erroneous models predictions, employing where possible computational algorithms such as GrowMatch (Kumar and Maranas, 2009) and Gapfilling (Kumar *et al*, 2007), to identify and correct the errors causing the incorrect viability predictions. We also analyzed the functions encoded by the intervals using knowledge from the literature, the SEED database (Overbeek *et al*, 2005), KEGG (Kanehisa *et al*, 2008), BSORF (http://bacillus.genome.jp), the annotation of the *B. subtilis* genome (Barbe *et al*, 2009), and the Subtiwiki (Lammers *et al*, 2010) to develop hypotheses for observed viability phenotypes that could be captured within the model. Through the course of this analysis, we encountered six distinct causes for *false positive* prediction errors and six distinct causes for *false negative* prediction errors (Table 2). Often, prediction errors were not a result of just one cause but a complex combination of several causes, and when some errors were corrected, new errors often emerged, highlighting new problems in the model or media formulations. The results of our model refinement effort, broken down by the 12 possible causes for prediction errors, are summarized in Table 2. The complete list of errors, the viability predictions affected by each error, and the action taken to correct the error are detailed in Supplementary Table S8.

The final corrected version of the *i*Bsu1103ΔtrpCD, named *i*Bsu1103V2ΔtrpCD, predicted deletion mutant viability with an accuracy of 99% in conditions where strains were viable, 78% in conditions where strains were not viable, and 96% overall (Table 1). Of the 14 knockout strains for which false positive predictions remain, we can attribute 6 to non-metabolic phenomena that cannot be simply captured within a steady-state metabolic model. The 8 remaining strains with prediction errors could not be corrected without introducing new prediction errors into the model. These conflicting errors indicate the presence of a regulatory mechanism controlling the expression of conflicting pathways. All

Christopher 1/10/12 11:58 PM
Deleted: 71

Christopher 1/10/12 11:58 PM
Deleted: 95

Christopher 1/11/12 12:09 AM
Deleted: 12

Christopher 1/11/12 12:00 AM
Deleted: remaining

Christopher 1/11/12 12:13 AM
Deleted: 10

Christopher 1/11/12 12:09 AM
Deleted: two

adjustments made to the *i*Bsu1103Δtrp were subsequently implemented in the *i*Bsu1103 model to create the *i*Bsu1103V2. The *i*Bsu1103V2 was then applied to the prediction of available wild-type growth phenotype data (i.e. Biolog (Oh *et al*, 2007), interval knockout data (Morimoto *et al*, 2008), and single gene inactivation (Kobayashi *et al*, 2003). This study revealed the *i*Bsu1103V2 model to be 93.4% accurate when predicting wild type growth phenotypes, a small increase compared to the 93.1% accuracy of the original *i*Bsu1103 model (Table 3).

Christopher 1/11/12 12:13 AM
Deleted: 92

Christopher 1/11/12 12:14 AM
Deleted: reduction

Christopher 1/11/12 12:14 AM
Deleted: from

Christopher 1/11/12 12:14 AM
Deleted: This reduction in accuracy is due to the reduced capacity by the *i*Bsu1103V2 model to predict Biolog growth conditions, which is the least reliable of the phenotype data used to validate the model.

Model-driven reinterpretation of experimental results

In the case of four deletion mutants (Δ0308, Δ0486, Δ0239, Δ0729), our efforts to reconcile model predictions with experimental observations inspired the reinterpretation of experiment results. Each of these deletion mutants is lacking the primary biosynthesis pathway for one or more amino acids or for nicotinate. As a result, the model predicted that none of these mutants should be viable in minimal media. Yet experimental results indicated that all four mutants were viable in minimal media. No reasonable adjustments could be identified to correct the conflicting prediction and explain the observed phenotype. Thus, we looked to the experiment itself for an alternative explanation. We determined that the initially observed growth of these strains was a result of cross-feeding of deletion mutants by the neighboring strains in the 96-well array deposited on the plate, which provided the needed compounds. Indeed, all four mutants failed to grow in individual MM plates and growth was restored by the addition of the needed compounds to the plate (Table 4, Supplementary information 5). This study demonstrates the importance of modeling as a means of identifying experimental results that have been misinterpreted.

Validation of model-driven viability hypotheses through rescue of deletion mutants

When the model predicts unviable strains, including after the correction of errors leading to *false positive* predictions, these predictions come bundled with one or more model-generated hypotheses about why the deletion mutant is not viable (e.g. the strain Δ0867 is not viable in NMS because it lacks heme biosynthesis pathways, and NMS does not contain heme). In these cases, we applied the model to identify compounds that could restore strain viability if added to the growth media (e.g. add heme to NMS to restore viability when lacking heme biosynthesis, see detailed examples in Supplementary information 6). We then verified the viability hypotheses experimentally by supplementing the media with the proposed metabolites (when commercially available) and testing for the growth of deletion mutants. When the model was adjusted in order to correct *false positive* predictions, this experimental validation was useful to confirm that these changes were correct.

When we applied this approach to the 18 mutants found to be unviable on minimal media (Table 4, Supplementary Table S4), the uncorrected model identified rescue metabolites for only 2 mutants. The refined model successfully identified rescue metabolites for another 11 mutants. Manual analysis of the knockout intervals revealed one additional rescue metabolite, malate, which restored viability to strain Δ0704 lacking a complete TCA cycle. However, no method could be found to capture this phenotype in the model. Similarly, strain Δ0730 lacking viability in minimal media was found to be missing genes encoding ATP synthase, but this phenotype could not be explained by the knockout of ATP synthase in the model without the introduction of additional regulatory constraints.

Similarly, we applied this approach to the 25 mutants found to be unviable in LB or NMS media (Table 5, Supplementary Table S4). The uncorrected model identified rescue metabolites for only three mutants, while the refined model successfully identified rescue metabolites for another 13 mutants. Manual analysis of knockout intervals revealed chorismate and shikimate as rescue metabolites of strains Δ0832

and Δ0915 respectively, but the model could not be modified to replicate these phenotypes. The original model correctly predicted the essentiality of the *ywpB* gene, but no rescue metabolite could be proposed to restore viability of this mutant. The corrected model identified FAD as a rescue for strain Δ0845, but the necessary reagents could not be acquired to test the prediction. Finally, no clear metabolic explanation could be produced to explain the loss of viability for the five remaining strains. However, given that four of these strains failed to grow on LB medium, non-metabolic explanations for these phenotypes appear very plausible.

Overall, the addition of metabolites restored the growth of 27 out of 42 deletion mutants, confirming our understanding of the metabolic pathways responsible for the loss of viability of these strains, and ruling out some pleiotropic effects of the deletion of multiple genes. Most vitally, these results serve to validate changes made to the metabolic model (Supplementary Table S5 and S6) and to the *in silico* formulation of LB (Supplementary Table S7) to improve prediction accuracy. For example, these experimental results inspired the removal of chorismate from LB and the addition of heme and shikimate to LB (Table 5, Supplementary information 6).

Discussion

In this work, we have established a comprehensive repertoire of the chromosome regions that are dispensable for *B. subtilis* growth in rich medium. This was achieved by individually deleting each chromosome region that did not include essential genes (Kobayashi *et al*, 2003) or genes involved in processes required to generate deletion mutants. We also avoided gene clusters encoding ribosomal RNAs due to their high redundancy. A total of 157 chromosome intervals were targeted for deletion, with sizes ranging from 2 to 159 kb. 146 of these intervals yielded viable mutant strains in rich media. The ten intervals which could not be deleted were split into smaller intervals enabling the identification of 4 individually essential genes and of 7 essential functions encoded by at least 2 genes. Our repertoire of deletions, which covers ~76% (3.22 Mb) of the chromosome, increases considerably the number of chromosome regions known to be dispensable, thus providing new possibilities to streamline the *B. subtilis* genome by combining deletions. Deleted intervals were replaced by a cassette that can be evicted from the chromosome without any scar (Supplementary information 1), enabling the iterative accumulation of deletions, as was done previously to assemble recombinant genomes (Itaya *et al*, 2008; Itaya *et al*, 2005), and to reduce the chromosome (Morimoto *et al*, 2008; Westers *et al*, 2003). We conclude that our repertoire of dispensable regions together with the corresponding collection of deletion mutant strains, which is publically available at the Genetic Strain Research Center [web link], represent a new resource that will facilitate genome-scale mapping of functional interactions and stimulate the progress of synthetic biology in *B. subtilis*.

pnoirot 12/30/11 3:12 PM
Comment: KOSEI will contact Hironori NIKI at Genetic Strain Research Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan for reference number and procedure to deposit deletion strains?

We further explored the phenotypes of our 146 deletion mutants by testing for viability in a rich chemically defined medium (NMS) and in two minimal media with distinct nitrogen sources (MM+Gln and MM+NH4). The growth phenotypes observed were compared with those predicted by the *i*Bsu1103 genome-scale metabolic model (Henry *et al*, 2009) of *B. subtilis 168*, revealing the original model to be

96% accurate for conditions where strains are viable, and 36% accurate for conditions where strains are not viable. We investigated the discrepancies between model predictions and observations, and we corrected errors in the model which led to these discrepancies, improving accuracy to 99% for viable strains and 71% for unviable strains. This work demonstrates the capacity of metabolic modeling to test the consistency of our understanding of the *B. subtilis* metabolism against our experimentally observed phenotypes, and to adapt our knowledge of *B. subtilis* when it is inconsistent with our observations. Based on this work, 19 new biosynthesis reactions and metabolite transport reactions were added to the model; constraints on directionality and reversibility were adjusted for 98 reactions to eliminate incorrect alternative pathways or to allow uptake of a metabolite; and incorrect associations linking metabolic genes to their cognate proteins and to the reactions they catalyze (called GPR associations) were removed or adjusted for 36 reactions, leading to changes in gene annotations (Table 2). In addition, in four instances, model predictions also helped to correct experimental errors caused by the cross-feeding of deletion mutants by neighbor colonies in plate assays (Table 4).

When our refined model predicted that a mutant strain would not be viable in a particular media condition, it also produced an explanation for why the viability was lost (e.g. growth in minimal media is lost because an amino acid biosynthesis pathway has been knocked out). In order to test the validity of these explanations, we applied our model to identifying how the growth medium might be supplemented to restore viability. For 27 out of 42 strains, our refined model successfully predicted media supplements to restore strain viability (Table 4, 5). This work not only further validates our metabolic model, but it also validates that the refinements made to the model to improve its predictive capacity were the correct refinements to make. This is important, because in many cases, there are multiple ways in which a model may be adjusted to correct a viability prediction. Our experimental validation helps to ensure that the most biologically relevant solution was selected.

Our combined modeling/experimental approach yielded a more complete and accurate annotation of the *B. subtilis* genome, an improved reconstruction of the metabolic pathways, and an enhanced model of metabolic behavior. We anticipate that our collection of deletion strains and our combined modeling/experimental approach could be extended to study additional growth conditions, allowing further reconstruction and refinement of the model in selected areas of the cell metabolism. However, this combined approach also faces some limitations from both the experimental and computational sides. Experimentally, the main limitation comes from the poor understanding of some observed metabolic phenotypes that prevent their incorporation into the model. For example, we found that YwfI (HemQ), which is required for heme biosynthesis (Dailey *et al*, 2010), is co-essential with YwfH, a reductase involved in the biosynthesis of the antibiotic bacilysin (Mahlstedt and Walsh, 2010). YwfI is annotated as a chlorite dismutase, and displays a catalase activity possibly involved in the elimination of endogenous hydrogen peroxide (Dailey *et al*, 2010). Thus, the synthetic lethality of the *ΔywfH ΔywfI* mutant, which is bypassed by the addition of heme in the medium, could reflect a role of YwfH in heme biosynthesis or its contribution to a yet uncharacterized YwfI-mediated detoxification pathway. Future studies will be necessary to clarify YwfI-YwfH interplay. In another example, the synthetic lethality of the *ΔaroA ΔccpA* mutant which can be rescued by the addition of shikimate to LB medium (Table 5), suggests that the *ΔaroA ΔccpA* mutant requires higher shikimate concentrations than the *ΔaroA* mutant. In the absence of CcpA, a key transcriptional regulator of central carbon metabolism (Fujita *et al*, 1995), the synthetic lethality likely arises from the deregulation of the chorismate biosynthesis or of the shikimate uptake. The metabolic model was adjusted to predict the correct phenotype of the *ΔaroA* mutant but the regulatory effects could not be taken into account.

Computationally, the main limitation is the current restriction of the model to include only mass balance and reversibility constraints, which prohibits the successful prediction of nonmetabolic phenotypes or even metabolic phenotypes that involve a regulatory component. In some cases, clearly metabolic phenotypes could not be explained by the model because the adjustment of the model to capture these phenotypes results in disruption of model predictions for other growth conditions and knockouts. This limitation could be partially overcome with the addition of regulatory constraints, such as those introduced in (Covert and Palsson, 2002).

In conclusion, the main outputs of this work are a repertoire of dispensable regions of the *B. subtilis* chromosome with the corresponding set of isogenic strains carrying single deletions, and a refined and more accurate metabolic model. In combination, these elements open new ways to generate biotechnologically useful strains by model-assisted streamlining of *B. subtilis* metabolism, and to further define the minimal genome that enables a bacterial cell to grow and divide.

Materials and methods

Media compositions

Bacillus subtilis strains were grown at 37°C in Luria Broth (LB) and in NMS, a rich medium of chemically defined composition. NMS is based on the minimal salts medium (10.8 g l^{-1} K_2HPO_4, 6 g l^{-1} of KH_2PO_4, 1 g l^{-1} of Na Citrate $2H_2O$, 2 g l^{-1} of K_2SO_4) supplemented with 0.4% glucose, 0.1% casamino acids (Difco Casamino Acids, Bacto), 0.01% L-tryptophane, 0.016% L-glutamine, 0.005% L-asparagine, 0.004% L-cystein, 0.01% L-histidine, trace elements (0.001 g l^{-1} of $MnCl_2 \cdot 4H_2O$, 0.0017 g l^{-1} of $ZnCl_2$, 0.00043 g l^{-1} of $CuCl_2 \cdot 2H_2O$, 0.0006 g l^{-1} of $CoCl_2 \cdot 6H_2O$ and 0.0006 g l^{-1} of $Na_2MoO_4 \cdot 2H_2O$), 0.1 mM of $FeCl_3$, 0.1 mM of $CaCl_2$, 1 mM of $MgSO_4$ and 1 mg l^{-1} of each following vitamins: B12, calcium pantothenate, nicotinic acid, pyridoxal, thiamine, folic acid, biotin and riboflavin. Solid media were obtained by adding 1.5% agar to the liquid NMS media. *B. subtilis* competent cells were transformed by the method of (Anagnostopoulos and Spizizen, 1961) modified according to (Jenkinson, 1983). The antibiotics phleomycin, neomycin, and chloramphenicol were added to NMS medium at final concentrations of 4, 15, and 5 µg ml^{-1}, respectively, and to LB medium at final concentrations of 8, 15, and 5 mg l^{-1}, respectively. Minimal salts medium supplemented with 0.5% glucose, 0.01% L-tryptophane and trace elements (MM) was used to assay nitrogen source utilization by adding either 2 g l^{-1} of ammonium sulfate (MM+NH_4) or 0.2% glutamine (MM+Gln) as the sole nitrogen source.

Construction of the deletion mutant strains

The deletion system is composed of the master strain in which all the deletions were introduced, and of a cassette allowing the positive selection of deletions and the eviction of the markers. The master strain is derived from the TF8A strain lacking 231 kb of the chromosome relative to wild type *B. subtilis* 168,

including the prophages SPβ and PBSX, and the prophage-like element *skin* (Westers *et al*, 2003). The *upp* gene of TF8A was replaced with a neomycin resistance gene under the control of the Lambda Pr promoter (λPr-*neo*), as described in Supplementary Figure S2, to give the master strain TF8A λPr-*neo::Δupp*. All deletions were introduced in the master strain by homologous replacement of the targeted chromosome region by a DNA fragment called "cassette *upp-phleo-cI*" (Figure 2, see Supplementary information 1), carrying the phleomycin-resistance gene for positive selection of cassette integration, and both the *upp* and *Psak-λcI* genes for counterselection and cassette eviction. The chromosome structure of each deletion mutant strain was verified by a PCR-based assay (Figure 2C, Supplementary information 1). All the primers used in this study are described in Supplementary Table S9.

Growth of deletion mutant strains

Growth of deletion mutant strains was assessed by the capacity to form colonies on plates. The master strain formed colonies of >2 mm diameter in 24 h on LB and on NMS. Three categories of growth were defined arbitrarily for deletion mutants based on visual inspection of the plates after 24 and 48 hours at 37°C. In the tables reporting experimental data (Table 4, 5, and Supplementary Table S2, S3 and S4), strain growth was annotated "+" if visible isolated colonies (diameter larger than 0.5 mm) appeared within 24 h and "slow" if visible colonies appeared within 48 h. Growth was annotated "-" if no colonies were obtained from the primary transformation, or if transformants were unable to form isolated colonies upon re-streaking on the same medium. Quantitative measurements of growth rates were performed in NMS medium using 96-well microtiter plates as described in Supplementary Information 3 (Supplementary Figure S3). Deletion mutant strains arrayed in the 96-well format were cultivated on LB plates at 37°C. For each strain, two independent colonies were assayed for growth on different nitrogen sources. Cells were picked and resuspended into 50 µL of MM using a multipronged device, and diluted

10 fold in MM. About 3 µL of the primary and diluted cell suspensions were deposited on MM+NH_4 and MM+Gln media. Plates were incubated at 37°C for up to 4 days, and each plates was photographed every 24h. Growth was scored as the capacity to form a patch within 96 h at 37°C. Lack of growth was systematically confirmed by streaking the strains grown on LB plates for single colony isolation on MM+NH_4 and MM+Gln plates.

Application of Flux Balance Analysis to Prediction/Reconciliation of Mutant Phenotypes

All deletions experimentally implemented in the course of this work were also simulated *in silico* using Flux Balance Analysis (FBA) (Orth *et al*, 2010) and a modified version of the *i*Bsu1103 genome-scale metabolic model of *B. subtilis* 168 (Henry *et al*, 2009). The *i*Bsu1103 model was altered to remove all reactions associated with the *trpC* and *trpD* genes, reflecting the mutation present in the mother strain utilized in all gene interval deletions. We refer to this modified model as *i*Bsu1103ΔtrpCD. FBA was then utilized with the *i*Bsu1103ΔtrpCD model to predict the viability of all gene interval knockout strains in LB, NMS, MM+NH_4, and MM+Gln media. In FBA, a set of linear constraints is established on the flux through each reaction involved in the metabolic pathways, representing the mass balance around each internal metabolite in the cell. Reaction fluxes are further constrained based on knowledge of the reversibility and directionality of the metabolic reactions, determined from thermodynamics (Henry *et al*, 2006; Jankowski *et al*, 2008). A linear optimization is then performed with these constraints such that the flux through the reaction representing cell growth (called the biomass reaction) is maximized subject to the mass balance constraints, the reversibility constraints, and the availability of nutrients in the media. Gene knockouts are then simulated by restricting the flux through metabolic reactions associated with the knocked out genes to zero. This then results in tighter restrictions on the conditions in which the biomass reaction can have a nonzero flux.

Acknowledgements

We are grateful to Marie-Françoise Noirot-Gros, Etienne Dervyn, and François Lecointe for stimulating discussions. This work was supported by the U.S. Department of Energy under contract DE-ACO2-06CH11357. The collection of *B. subtilis* deletion mutant strains has been deposited at the Genetic Strain Research Center (Japan) under accession numbers xxxx

pnoirot 12/29/11 1:03 AM
Comment: KOSEI will contact H. Niki to determine the procedure to do this.

The authors declare that they have no conflict of interest.

Author Contributions:

pnoirot 12/29/11 1:03 AM
Comment: ALL AUTHORS PLEASE CHECK

KT performed all the deletion mutant strains and phenotypic analyses. CH performed computational analyses and refined the metabolic model. JFZ, MPC, FX developed computational tools for interval design. EJ and V.B developed genetic tools for deletion making. SDE, RLS and PN conceived the project. KT, CH, JFZ and PN designed the study and analyzed the results. CH and PN wrote the manuscript.

References

Anagnostopoulos C, Spizizen J (1961) Requirements for transformation in Bacillus subtilis. *Journal of bacteriology* **81:** 741-746.

Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Medigue C, Danchin A (2009) From a consortium sequence to a unified sequence: the Bacillus subtilis 168 reference genome a decade later. *Microbiology* **155:** 1758-1775.

Covert MW, Palsson BO (2002) Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *J Biol Chem* **277:** 28058-28064.

Dailey TA, Boynton TO, Albetel AN, Gerdes S, Johnson MK, Dailey HA (2010) Discovery and Characterization of HemQ: an essential heme biosynthetic pathway component. *J Biol Chem* **285:** 25978-25986.

de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC (2003) Perspective: Evolution and detection of genetic robustness. *Evolution* **57:** 1959-1972.

Fabret C, Ehrlich SD, Noirot P (2002) A new mutation delivery system for genome-scale approaches in Bacillus subtilis. *Mol Microbiol* **46:** 25-36.

Fischer E, Sauer U (2005) Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. *Nat Genet* **37:** 636-640.

Fujita Y, Miwa Y, Galinier A, Deutscher J (1995) Specific recognition of the *Bacillus subtilis gnt cis*-acting catabolite-responsive element by a protein complex formed between CcpA and seryl-phosphorylated HPr. *Mol Microbiol* **17:** 953-960.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:** 52-56.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103:** 425-430.

Glass JI, Hutchison CA, 3rd, Smith HO, Venter JC (2009) A systems biology tour de force for a near-minimal bacterium. *Mol Syst Biol* **5:** 330.

Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326:** 1268-1271.

Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato J (2005) Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. *Mol Microbiol* **55:** 137-149.

Henry C, Overbeek R, Stevens RL (2010) Building the blueprint of life. *Biotechnol J* **5:** 695-704.

Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V (2006) Genome-scale thermodynamic analysis of Escherichia coli metabolism. *Biophys J* **90:** 1453-1461.

Henry CS, Zinner JF, Cohoon MP, Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biol* **10:** R69.

Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, Raineri E, Garriga-Canut M, Serrano L (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452:** 840-845.

Itaya M, Fujita K, Kuroki A, Tsuge K (2008) Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat Methods* **5:** 41-43.

Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc Natl Acad Sci U S A* **102:** 15971-15976.

Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys J* **95:** 1487-1499.

Jenkinson HF (1983) Altered arrangement of proteins in the spore coat of a germination mutant of Bacillus subtilis. *J Gen Microbiol* **129:** 1945-1958.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36:** D480-484.

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, et al. (2003) Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A* **100:** 4678-4683.

Kolisnychenko V, Plunkett G, 3rd, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G (2002) Engineering a reduced Escherichia coli genome. *Genome Res* **12:** 640-647.

Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castano-Diez D, Chen WH, Devos D, Guell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, et al. (2009) Proteome organization in a genome-reduced bacterium. *Science* **326:** 1235-1240.

Kumar VS, Dasika MS, Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8:** 212.

Kumar VS, Maranas CD (2009) GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput Biol* **5:** e1000308.

Lammers CR, Florez LA, Schmeisky AG, Roppel SF, Mader U, Hamoen L, Stulke J (2010) Connecting parts with processes: SubtiWiki and SubtiPathways integrate gene and pathway annotation for Bacillus subtilis. *Microbiology* **156:** 849-859.

Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC (2009) Metabolic engineering of a reduced-genome strain of Escherichia coli for L-threonine production. *Microb Cell Fact* **8:** 2.

Mahlstedt SA, Walsh CT (2010) Investigation of anticapsin biosynthesis reveals a four-enzyme pathway to tetrahydrotyrosine in Bacillus subtilis. *Biochemistry* **49:** 912-923.

Morimoto T, Kadoya R, Endo K, Tohata M, Sawada K, Liu S, Ozawa T, Kodama T, Kakeshita H, Kageyama Y, Manabe K, Kanaya S, Ara K, Ozaki K, Ogasawara N (2008) Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis. *DNA Res* **15:** 73-81.

Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* **282:** 28791-28799.

Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* **28:** 245-248.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome Escherichia coli. *Science* **312:** 1044-1046.

Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **10:** 410-422.

Serrano L (2007) Synthetic biology: promises and challenges. *Mol Syst Biol* **3:** 158.

Sierro N, Makita Y, de Hoon M, Nakai K (2008) DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res* **36:** D93-96.

Stelling J, Sauer U, Szallasi Z, Doyle FJ, 3rd, Doyle J (2004) Robustness of cellular functions. *Cell* **118:** 675-685.

Westers H, Dorenbos R, van Dijl JM, Kabel J, Flanagan T, Devine KM, Jude F, Seror SJ, Beekman AC, Darmon E, Eschevins C, de Jong A, Bron S, Kuipers OP, Albertini AM, Antelmann H, Hecker M, Zamboni N, Sauer U, Bruand C, et al. (2003) Genome engineering reveals large dispensable regions in Bacillus subtilis. *Mol Biol Evol* **20:** 2076-2090.

Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS, Kim SC (2002) Minimization of the Escherichia coli genome using a Tn5-targeted Cre/loxP excision system. *Nat Biotechnol* **20:** 1018-1023.

Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH, Wodke JA, Guell M, Martinez S, Bourgeois R, Kuhner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutierrez-Gallego R, Russell RB, Gavin AC, Bork P, et al. (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326:** 1263-1268.

Figure legends

Figure 1. Chromosome interval deletion maps: results from experiments and from model predictions

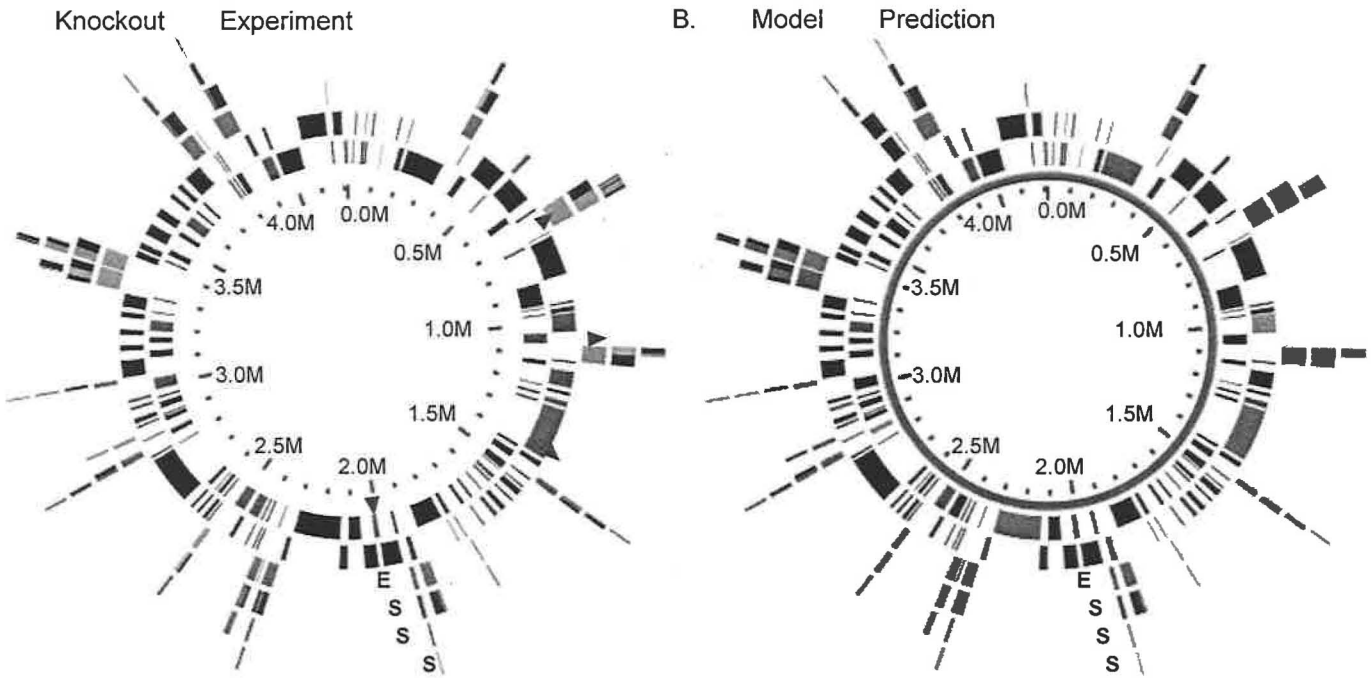
Chromosome coordinates are indicated in the inner circle. The 156 intervals targeted for deletion are represented on the successive outer circles as follows: the 136 intervals dispensable for growth on NMS are alternated on the second and third circles; the 21 intervals essential for growth on NMS are on the 4th circle (Ess), and the smaller intervals resulting from the iterative splitting of essential intervals are shown on the outer circles (S1, S2, S3). **(A)** Experimental results from interval deletions are color-coded as follows: essential for growth on LB and NMS (red); on NMS only (orange); on minimal medium only (green); and dispensable on all tested media (black). Among the intervals essential for growth on minimal medium, those required for the utilization of ammonium (purple arrowheads) and glutamine (brown diamond) as sole nitrogen source are indicated. **(B)** Results from model predictions are color-coded as follows: correct prediction with *i*Bsu1103 (black); correct prediction with refined model *i*Bsu1103v2 (green); false positive (red); false negative (orange).

Figure 2. Systematic interval deletion, chromosome structure verification and qualitative phenotyping of colony growth

(A) A linear DNA molecule containing the *upp-phleo-cI* cassette (red) flanked by 1.5 kb segments homologous to the chromosome regions to be kept (grey shading) is generated by a joining PCR reaction. Upon transformation of competent *B. subtilis* cells for phleomycin-resistance (Phleo^{R}), the integration of the cassette by double crossing-over into the chromosome replaces the interval targeted for deletion (multicolored). **(B)** The replacement of a dispensable interval by the cassette yields a deleted chromosome structure. When an interval encodes a function essential for cell survival, its deletion and

replacement is impossible. However, some spontaneous Phleo^R mutants can arise, and in some rare instances, the parental and deleted chromosomes can co-exist in the same cell (merodiploidy) to form a viable strain. Therefore, the chromosome structures of potential deletion mutants are checked systematically by a PCR assay. Primers (blue arrows) are positioned on the chromosome structures. The phleo3-phleo5, p1-p2 and p3-p4 primer pairs are used to generate the transforming DNA molecule (in A). The primer combinations phleo3-cI2RV and p5-p6 are used to verify deleted and parental chromosome structures, respectively. **(C)** The typical results of the structure verification of 7 candidate deletion mutants are shown. PCR assay were performed on isolated single colonies, and parental (Par) and deleted (Del) chromosome structures were revealed after electrophoretic separation of PCR products of diagnostic sizes (arrows). Clone n°2 is a spontaneous Phleo^R mutant whereas the 6 remaining clones have the expected interval deletion. C is a positive control for PCR and M are DNA size markers. **(D)** The deletion mutants were categorized based on the size of isolated colonies. The deletion mutant strains were streaked on fresh selective plate and incubated at 37°C, and the colony size was measured after 24 and 48h. The strains forming colonies with a diameter larger than 0.5 mm within 24 h were categorized as "positive". The strains forming colonies with a diameter less than 0.5 mm at 24 h and larger than 0.5 mm within 48 h were categorized as "slow". The strains unable to form colonies larger than the background growth of sensitive strains (0.5 mm within 48h) were categorized as "negative".

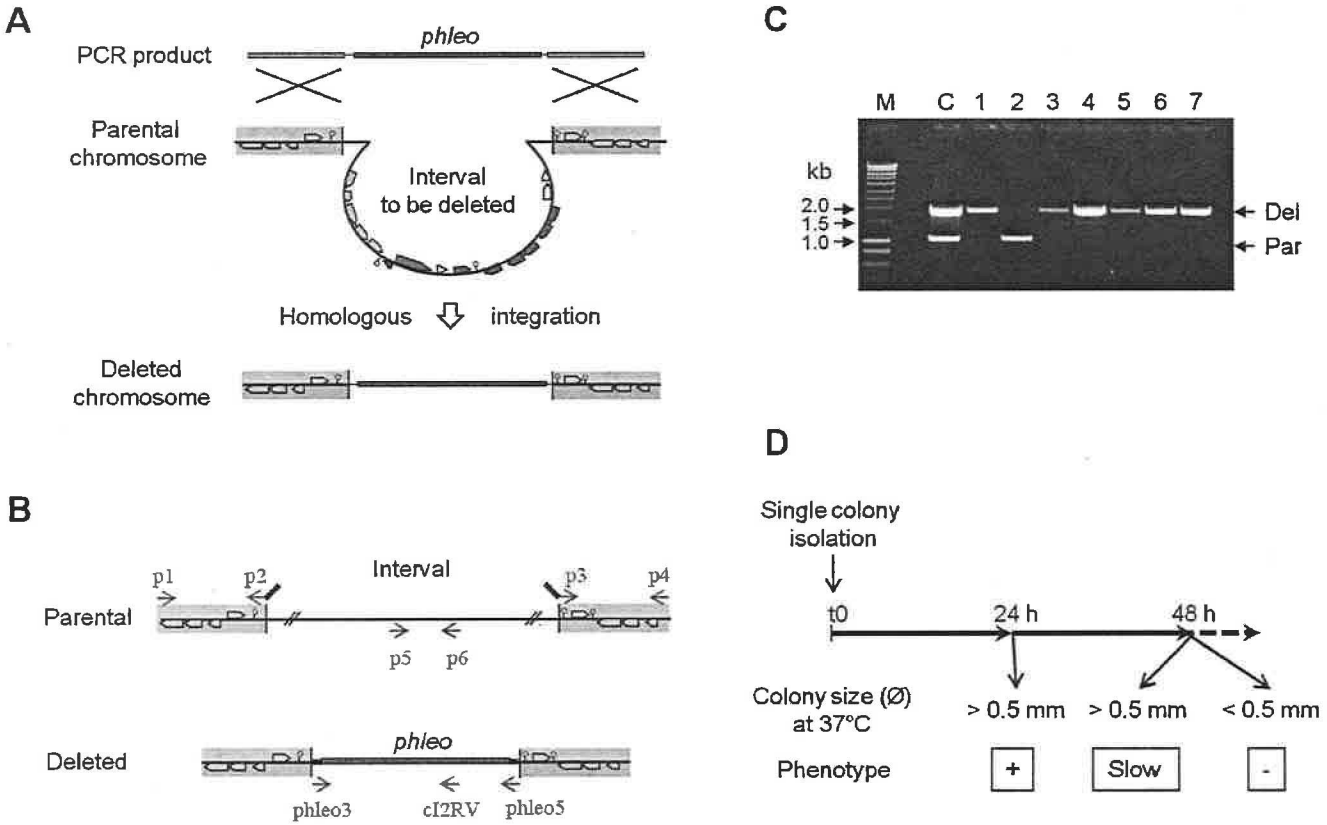
Figure 1



Essential on LB ( ), NMS ( ), MM ( ), and dispensable ( ).

Model prediction: correct ( ), corrected ( ),

Figure 2


A
PCR product
phleo
Parental chromosome
Interval to be deleted
Homologous integration
Deleted chromosome
B
Parental
p1
p2
Interval
p3
p4
p5
p6
Deleted
phleo
phleo3
cI2RV
phleo5
C
M C 1 2 3 4 5 6 7
kb
2.0
1.5
1.0
Del
Par
D
Single colony isolation
t0
24 h
48 h
Colony size (Ø) at 37°C
> 0.5 mm
> 0.5 mm
< 0.5 mm
Phenotype
+
Slow
-

Tables

Table 1. Accuracy of original and refined models in predicting deletion strain viability

Phenotype	Experimental observations	Correct *i*Bsu1103ΔtrpCD	Correct *i*Bsu1103V2ΔtrpCD
LB+	146	140 (96%)	146 (99%)
LB-	10	1 (10%)	4 (40%)
NMS+	135	130 (96%)	134 (99%)
NMS-	22	8 (36%)	17 (77%)
MM_NH4+	111	107 (96%)	110 (99%)
MM_NH4-	24	10 (42%)	22 (88%)
MM_Gln+	115	111 (97%)	114 (99%)
MM_Gln-	20	8 (40%)	16 (75.0%)
Overall+	507	488 (96%)	504 (99%)
Overall-	76	27 (35%)	59 (78%)
Overall	583	515 (88%)	563 (96%)

The viability of deletion mutant strains was experimentally observed and computationally predicted on LB, NMS, and on minimal medium containing ammonium (MM_NH4) and glutamine (MM_Gln) as nitrogen source. Phenotypes were divided into strains that were observed to be viable (labeled with a "+") and not viable (labeled with a "-") in the experiments. This division highlights the significant differences in prediction accuracy for viable versus nonviable strains. The number of strains that fit each phenotypes in our experiments (second column), the fraction of strains that were correctly predicted by the initial *i*Bsu1103ΔtrpCD (third column) and the refined *i*Bsu1103V2ΔtrpCD model (fourth column) are indicated.

Christopher 1/10/12 11:46 PM
Formatted ... [1]
Christopher 1/10/12 11:46 PM
Formatted Table ... [2]
Christopher 1/7/12 1:48 PM
Deleted: 145 (99%)
Christopher 1/10/12 11:46 PM
Formatted ... [3]
Christopher 1/10/12 11:50 PM
Deleted: 11... ... [4]
Christopher 1/10/12 11:50 PM
Formatted ... [5]
Christopher 1/7/12 1:48 PM
Deleted: 5 (45%)
Christopher 1/10/12 11:46 PM
Formatted ... [6]
Christopher 1/10/12 11:46 PM
Formatted Table ... [7]
Christopher 1/7/12 1:48 PM
Deleted: 133 (99%)
Christopher 1/10/12 11:46 PM
Formatted ... [8]
Christopher 1/10/12 11:46 PM
Formatted ... [9]
Christopher 1/10/12 11:46 PM
Formatted ... [10]
Christopher 1/10/12 11:46 PM
Formatted Table ... [11]
Christopher 1/7/12 1:48 PM
Deleted: 18 (75%)
Christopher 1/10/12 11:46 PM
Formatted ... [12]
Christopher 1/10/12 11:46 PM
Formatted ... [13]
Christopher 1/10/12 11:46 PM
Formatted Table ... [14]
Christopher 1/7/12 1:48 PM
Deleted: 15 (75%)
Christopher 1/10/12 11:46 PM
Formatted ... [15]
Christopher 1/7/12 1:48 PM
Deleted: 502 (99%)
Christopher 1/10/12 11:46 PM
Formatted ... [16]
Christopher 1/10/12 11:51 PM
Deleted: 77... ... [17]
Christopher 1/10/12 11:46 PM
Formatted ... [18]
Christopher 1/10/12 11:53 PM
Deleted: 584... ... [19]
Christopher 1/10/12 11:53 PM
Formatted ... [20]

Table 2. Model corrections, intervals, and phenotypes associated with each type of error

Class of error	Change made to model	Associated strains	Associated phenotypes
FP: Missing metabolites in biomass	Added 6 compounds to biomass	9	23
FN: Extra metabolites in biomass	None	0	0
FP: Incorrect reaction GPR	GPR adjusted on 2 reactions	2	5
FN: Incorrect reaction GPR	GPR adjusted on 4 reactions	5	8
FP: Incorrect isozymes in GPR	10 isozymes removed from 8 reactions	12	22
FN: Isozymes missing from GPR	4 isozymes added to 19 reactions	5	10
FP: Extra pathways in model	2 reaction removed	2	2
FN: Missing pathways in model	12 reactions added	23	36
FP: Under-constrained reversibility	Made 9 reactions irreversible	10	14
FN: Over-constrained reversibility	Made 4 reactions reversible	8	11
FP: Extra nutrients in media	Removed 1 compound from LB and 1 compound from NMS	2	2
FN: Missing nutrients in media	Added 9 compounds to LB and 1 compounds to NMS	46	47

Christopher 1/9/12 1:01 AM
Formatted: Left

Christopher 1/9/12 1:01 AM
Formatted Table

Christopher 1/9/12 1:01 AM
Deleted: 25

Christopher 1/9/12 1:01 AM
Formatted Table

Christopher 1/9/12 1:32 AM
Deleted: 1 ... ... [21]

Christopher 1/9/12 1:32 AM
Deleted: 3 ... ... [22]

Christopher 1/9/12 1:27 AM
Deleted: 1...12... ... [23]

Christopher 1/9/12 1:30 AM
Deleted: 8 ...20 ... ... [24]

Christopher 1/9/12 1:03 AM
Deleted: 5 ... ... [25]

Christopher 1/9/12 1:04 AM
Deleted: 19 ... ... [26]

Christopher 1/9/12 1:18 AM
Deleted: 78 ... ... [27]

Christopher 1/9/12 1:23 AM
Deleted: 20 ... ... [28]

Christopher 1/9/12 1:02 AM
Deleted: 14 ...2 ... ... [29]

The twelve classes of model errors that led to incorrect strain viability predictions are described. Each class of error is associated with "false positive" (FP) or "false negative" (FN) predictions. FP indicates a condition where the model predicted that an unviable strain would grow; FN indicates a condition where the model predicted that a viable strain would not grow. The changes made to the model to correct the error (column 2), the numbers of mutant strains (column 3) and of strain phenotypes corrected by the model changes (column 4) are indicated. The complete list of errors, the viability predictions affected by each error, and the correction of errors are detailed in Supplementary Table S8.

Table 3. Accuracy of the original and refined models in predicting wild type phenotypes

	*i*Bsu1103	*i*Bsu1103V2
Morimoto KO in LB media	58/63 (92.1%)	59/63 (93.7%)
Morimoto KO in MM media	54/63 (85.7%)	56/63 (88.9%)
Kobayashi 271 essential genes	195/215 (90.7%)	192/215 (89.3%)
Kobayashi nonessential genes	872/888 (98.2%)	870/882 (98.3%)
New 274 essential genes	196/218 (89.9%)	195/218 (89.4%)
New nonessential genes	870/885 (98.3%)	870/879 (98.6%)
Biolog conditions	218/271 (80.4%)	216/271 (79.7%)
Overall accuracy	**1396/1500 (93.1%)**	1396/1494 (93.4%)

The accuracy of the original and refined *i*Bsu1103 models in predicting gene essentiality, Biolog growth conditions, and interval knockout viability for the wild-type *B. subtilis* 168 strain. The type of experimental phenotypic data is indicated in the first column. In the next columns, the ratio between the number of growth phenotypes correctly predicted and the number of phenotypes for which a prediction was possible is indicated for the original and refined models, respectively. Prediction accuracy is in parentheses. Gene essentiality datasets comprise the previously published 271 essential genes (Kobayashi *et al*, 2003) and the new updated set of 274 essential genes highlighted in this work.

Christopher 1/7/12 1:50 PM
Formatted Table

Christopher 1/7/12 1:50 PM
Deleted: 58/63 (92.1%)

Christopher 1/7/12 1:50 PM
Deleted: 56/63 (88.9%)

Christopher 1/7/12 1:50 PM
Deleted: 192/215 (89.3%)

Christopher 1/7/12 1:50 PM
Deleted: 867/884 (98.1%)

Christopher 1/7/12 1:50 PM
Deleted: 195/218 (89.4%)

Christopher 1/7/12 1:50 PM
Deleted: 867/881 (98.4%)

Christopher 1/7/12 1:50 PM
Deleted: 207/271 (76.4%)

Christopher 1/7/12 1:50 PM
Deleted: 1383/1496 (92.4%)

Table 4. Reconciliation of experiments and model predictions: rescue of growth on minimal media

Int.	Viability $NMS/MM_{Gln}/MM_{NH4}$	Prediction $NMS/MM_{Gln}/MM_{NH4}$	Function lost by KO	MM media supplement	Model Refinement
0308	+/+/+ → +/-/-	+/-/-	ile/val synthesis	Ile+val (+)	none, experiment error
0486	+/+/+ → +/-/-	-/-/-→+/-/-	ile/val synthesis	Ile+val (+)	adjusted GPR for NMS, experiment error for MM
0239	+/+/+ → +/-/-	+/-/-	phe synthesis nad synthesis	Phe+nicotinate (+)	none, experiment error
0729	+/+/+ → +/-/-	+/+/+	ile/val synthesis	ile (+) 2oxo-butanoate (+)	experiment error, no model refinement found
0620	+/-/-	+/+/+ → +/-/-	thiamine synthesis	thiamine (+)	TPP path added, TPP added biomass, adjusted GPR
0235	+/-/-	+/+/+ → +/-/-	thiamine synthesis	thiamine (+)	TPP path added, TPP added biomass, adjusted GPR
0642	+/-/-	+/+/+ → +/-/-	pyridoxal synthesis	pyridoxal (+)	pyridoxal path added, pyridoxal added biomass
0644	+/-/-	+/+/+ → +/-/-	cysteine synthesis	cyst (+)	adjusted reversibility
0681	+/-/-	+/+/+ → +/-/-	cysteine synthesis	cyst (+)	adjusted GPR
0281	+/+/-	+/+/+ → +/+/-	aspartate synthesis	asp (+)	adjusted reversibility
0704	+/+/-	+/+/+	TCA cycle	malate (+)	none found
0698	slow/-/-	+/+/+ → +/-/-	ile/val/leu synthesis	Ile+val+leu (-)	removed reactions
0546	+/-/-	+/+/+ → +/-/-	ser synthesis tyr/phe synthesis	ser+tyr+phe (+)	adjusted reversibility
0730	+/-/-	+/+/+	ATP synthase	none	none found
0608	+/-/-	+/+/+	unknown	none	none found
0661	+/-/-	+/+/+ → -/-/-	folate synthesis	folate (+)	adjusted GPR, but path in NMS remains unknown
0581	+/+/-	+/+/+	unknown	none	none found
0641	+/-/+	+/+/+	unknown	none	none found

Christopher 1/7/12 2:02 PM
Deleted: →+/-/-

Christopher 1/7/12 2:03 PM
Deleted: →+/-/-

Christopher 1/7/12 2:03 PM
Deleted: reactions removed and added

The viability of deletion mutant strains was experimentally observed and computationally predicted on 3 distinct chemically defined media. Analysis of *i*Bsu1103Δtrpsuggested missing potential functions that could explain observed/predicted discrepancies in minimal media. Metabolites added to MM media are listed and resulting growth is indicated in parentheses. The types of changes made in the model to accommodate the observations are listed and the resulting changes in predictions are pointed by →.

Table 5. Reconciliation of experiments and model predictions: rescue of growth on complex (LB) and chemically defined (NMS) rich media

Int.	Viability LB/NMS	Prediction LB/NMS	Function lost by KO	Medium supplement	Model Refinement
0091	slow/-	+/+ →+/-	FolEA, MtrB	folate (-) [a]	adjusted GPR
0291	+/-	+/+ →+/-	Pantothenate synthesis	pantothenate (+) [a]	pantothenate removed from NMS
0161	+/-	+/-	Purine synthesis	adenine (+) [a]	none needed
0720	slow/-	+/+ → +/-	Ribose-5P epimerase	ribose (+) [a]	adjusted reversibility
0853	+/-	+/- → +/-	Shikimate kinase AroK	chorismate (-) [a, c]	adjusted *in silico* LB composition,added transporter
0872	+/-	+/+ → +/+?	GMP synthesis	guanosine (+) [a]	adjusted *in silico* LB composition
0897	+/-	+/-	purine synthesis	DNA (+) [a]	none needed
0867	slow/-	+/+ → +/-	Heme synthesis	hemin (+) [a]	heme added to biomass and LB composition
0896	+/-	+/+ → +/-	Putative protein lipoate ligase	lipoate (-) [a]	none
0895	slow/-	+/+ → +/-	Heme synthesis	hemin (+) [a]	heme added to biomass and LB composition
0849	+/-	+/+ → +/-	Lipoate synthesis	lipoate (+) [a]	lyposyl-protein added to biomass and lipoate to LB composition
0857	+/-	+/+	Glycine cleavage system	n.t.	none found
0911	-/-	+/+ → -/-	Peptidoglycan synthesis	LL-2,6-diamino-pimelate (-) [b]	adjusted GPR
0845	-/-	+/+ → -/-	FAD synthesis	n.t.	FAD added to biomass
0787-3	-/-	+/+	Co-essential gene pair *glnR glnA*	n.t.	no purely metabolic explanation
0883	-/-	+/+	Co-essential gene pair *yneE yneF*	n.t.	no purely metabolic explanation
0910	-/-	+/+	Essential gene *rnpB*	n.t.	no purely metabolic explanation
0832	-/-	+/+	Co-essential gene pair *tyrA hisC*	chorismate (+) [b]	none found
0833	-/-	+/- → -/-	Chorismate mutase and Shikimate pathway	chorismate (+) [b]	adjusted *in silico* LB composition, adjusted GPR

Christopher 1/7/12 2:26 PM
Deleted: +/-

Christopher 1/7/12 2:27 PM
Deleted: -

0797	-/-	+/+	Co-essential region	n.t.	no purely metabolic explanation
0915	-/-	+/- → +/-	Co-essential gene pair *ccpA aroA*	shikimate (+) [a,b]	no purely metabolic explanation
0914	slow/-	+/-	Chorismate mutase AroA	shikimate (+) [a,b]	none needed
0906	-/-	-/-	Essential gene *ywpB*	n.t.	none needed
0825	-/-	+/+ → +/-	Co-essential gene pair *ywfI ywfH*	hemin (+) [a,b]	adjusted GPR, heme added to biomass and LB composition
0903	slow/-	+/+ → +/-	heme synthesis gene *ywfI*	hemin (+) [a,b]	

The viability of deletion mutant strains was experimentally observed and computationally predicted on complex (LB) and chemically defined (NMS) media. Analysis of *i*Bsu1103Δtrp and of the available knowledge suggested missing potential metabolic functions that could explain observed lack of growth on one medium. Metabolites added to the medium are listed and resulting growth is indicated in parentheses. The types of changes made in the model to accommodate the observations are listed and the resulting changes in predictions are pointed by→. a, NMS is supplemented. b, LB is supplemented. c, chorismate is highly unstable in NMS medium.

Page 32: [1] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [2] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [3] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [4] Deleted	Christopher	1/10/12 11:50 PM
11		
Page 32: [4] Deleted	Christopher	1/10/12 11:50 PM
2		
Page 32: [4] Deleted	Christopher	1/10/12 11:50 PM
18		
Page 32: [5] Formatted	Christopher	1/10/12 11:50 PM
Left		
Page 32: [6] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [7] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [8] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [9] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [10] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [11] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [12] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [13] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [14] Change	Christopher	1/10/12 11:46 PM
Formatted Table		
Page 32: [15] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [16] Formatted	Christopher	1/10/12 11:46 PM
Left		
Page 32: [17] Deleted	Christopher	1/10/12 11:51 PM
77		
Page 32: [17] Deleted	Christopher	1/10/12 11:51 PM
28		

Page 32: [17] Deleted Christopher 1/10/12 11:52 PM

36

Page 32: [17] Deleted Christopher 1/7/12 1:48 PM

55 (71%)

Page 32: [18] Formatted Christopher 1/10/12 11:46 PM

Left

Page 32: [19] Deleted Christopher 1/10/12 11:53 PM

584

Page 32: [19] Deleted Christopher 1/10/12 11:53 PM

516

Page 32: [19] Deleted Christopher 1/7/12 1:48 PM

557 (95%)

Page 32: [20] Formatted Christopher 1/10/12 11:53 PM

Font:

Page 33: [21] Deleted Christopher 1/9/12 1:32 AM

1

Page 33: [21] Deleted Christopher 1/9/12 1:33 AM

1

Page 33: [21] Deleted Christopher 1/9/12 1:33 AM

3

Page 33: [22] Deleted Christopher 1/9/12 1:32 AM

3

Page 33: [22] Deleted Christopher 1/9/12 1:34 AM

3

Page 33: [22] Deleted Christopher 1/9/12 1:34 AM

6

Page 33: [23] Deleted Christopher 1/9/12 1:27 AM

1

Page 33: [23] Deleted Christopher 1/9/12 1:27 AM

12

Page 33: [23] Deleted Christopher 1/9/12 1:29 AM

8

Page 33: [23] Deleted Christopher 1/9/12 1:29 AM

17

Page 33: [24] Deleted Christopher 1/9/12 1:30 AM

8

Page 33: [24] Deleted Christopher 1/9/12 1:30 AM

20

Page 33: [24] Deleted	Christopher	1/9/12 1:30 AM

4

Page 33: [24] Deleted	Christopher	1/9/12 1:31 AM

41

Page 33: [25] Deleted	Christopher	1/9/12 1:03 AM

5

Page 33: [25] Deleted	Christopher	1/9/12 1:03 AM

3

Page 33: [25] Deleted	Christopher	1/9/12 1:12 AM

4

Page 33: [26] Deleted	Christopher	1/9/12 1:04 AM

19

Page 33: [26] Deleted	Christopher	1/9/12 1:04 AM

6

Page 33: [26] Deleted	Christopher	1/9/12 1:04 AM

17

Page 33: [27] Deleted	Christopher	1/9/12 1:18 AM

78

Page 33: [27] Deleted	Christopher	1/9/12 1:19 AM

5

Page 33: [27] Deleted	Christopher	1/9/12 1:22 AM

9

Page 33: [28] Deleted	Christopher	1/9/12 1:23 AM

20

Page 33: [28] Deleted	Christopher	1/9/12 1:25 AM

2

Page 33: [28] Deleted	Christopher	1/9/12 1:26 AM

5

Page 33: [29] Deleted	Christopher	1/9/12 1:02 AM

14

Page 33: [29] Deleted	Christopher	1/9/12 1:02 AM

2

Page 33: [29] Deleted	Christopher	1/9/12 1:02 AM

12

Page 33: [29] Deleted	Christopher	1/9/12 1:02 AM